

# Weekly Report

Lu Junhua

2016 年 6 月 26 日

## Done

- Wrote a list on the problems addressed in the review. Discussed about the revision of VIS paper. We were assigned with different tasks, and we would cooperate during the following weeks.
- Finished course essays. Took an exam. Prepared for next week's course speech.

## To do

- A course speech on Tuesday.
- After that, begin revise the model and related part of VIS paper.

## 论文

- 最近作业较忙, 加上周末发高烧现在还没好, 看的不多, 下周会继续努力. 一篇是我们要报告的文章, 一篇是Zhenhui Li提到的文章, 一篇是可视化领域的.
- *Non-negative Matrix Factorization with Sinkhorn Distance* 本文讲了一种非负矩阵分解算法. 矩阵分解在各地领域都有广泛运用, 这篇文章提出的方法是有两个特点的: 一方面考虑到本身特征的相关性, 一方面又要考虑保持几何信息(流形上的). 为此他们使用了一个叫做推土机距离的东西来度量矩阵的差异, 并用最临近图约束几何信息. 这篇文章给我的收获在于, 有很多新名词出现了, 比如KL散度, 推土机距离, 这些我以前没见过的度量指标, 再加上临近图这样的术语. 我觉得在后面我们组的项目可能也用到这些术语和指标. 其中推土机距离是运输问题的一个异化版本, 由于限制比较严格, 也采用了其他方法来逼近这个解. 最后本文提出的新算法应用于图像聚类 and 纹理识别都取得了很好的效果.
- *Crime Rate Inference with Big Data* Zhenhui Li来报告时候最后提到的文章. 当时只讲了用了什么数据/为什么用, 一些小问题诸如为什么用了demographics数据再用geographical数据不会有显著提升. 但是她没有具体说预测到哪一个粒度以及用了什么方法. 看了下原文, 其实就是用了比较简单的线性回归方法和一个稍微复杂的negative binomial regression. 从模型本身来看确实后者效果会好一些, 实际运行结果也是如此. 值得注意的是文中其实做了很多简单的可视化工作, 如果加上特征选择之类的, 完全可以做个小的可视分析系统. 当然从另一个角度说, 作者做的确实很仔细, 在特征选择, 相关性, 显著性检验还有特征提取方面做的比较到位的.
- *UTOPIAN: User-Driven Topic Modeling Based on Interactive Nonnegative Matrix Factorization* 这是一篇13年的VIS文章, 由于我们模型被reviewer抨击写的太复杂看不懂, 最近也想看看人家怎么写

的. 这篇文章也是一个Non-negative Matrix Factorization的可视化, 我看了其中的交互几乎完全依赖于这个模型本身的强大功能. 其模型用数学语言概括其实不复杂, 不想我们的模型有好几层. 有一个值得学习的地方那就是把变量名用表格列出来了, 这一点我们可以借鉴, 尽管可能是一个非常长的表格, 但可能会让读者好受一点. 此外文中提到的t-SNE方法也是和MDS等类似的降维可视化方法, 作者对该方法改进以后可以交互的让点布局变得更直观, 减少视觉错乱.